



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2011

Datengeleitete Korpuspragmatik: Korpusvergleich als Methode der Stilanalyse

Scharloth, Joachim ; Bubenhofer, Noah

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-111285>

Book Section

Published Version

Originally published at:

Scharloth, Joachim; Bubenhofer, Noah (2011). Datengeleitete Korpuspragmatik: Korpusvergleich als Methode der Stilanalyse. In: Felder, Ekkehard; Müller, Marcus; Vogel, Friedemann. Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen. Berlin, New York: Walter de Gruyter GmbH, 195-230.

Datengeleitete Korpuspragmatik

Korpusvergleich als Methode der Stilanalyse

Joachim Scharloth und Noah Bubenhofer

Inhalt

1. Korpuslinguistik und Korpuspragmatik
 - 1.1 Korpuspragmatik: Korpuslinguistik als Methode einer kultur- und sozialwissenschaftlich interessierten Linguistik
 - 1.2 Korpuslinguistik als Denkstil: corpus based vs. corpus driven
 - 1.3 Datengeleitete Korpusanalyse und Korpuspragmatik
 2. Zum Stilbegriff
 - 2.1 Das Stilkonzept der Lebensstilforschung
 - 2.2 Stilkonzepte in der Linguistik
 - 2.3 Korpuslinguistische Operationalisierung des Stilbegriffs
 3. Die semtracks-Stilanalyse
 - 3.1 Komplexe n-Gramm-Analyse
 - 3.2 Datengeleitete Findung relevanter Kategorien auf der Tokenebene
 - 4.1 Kriterien bei der Korpuserstellung
 - 4.2 Das GerMov-Korpus und seine Subkorpora
 - 4.3 Annotation
 5. Kommunikationsstile in korpuslinguistischer Perspektive
 - 5.1 Berechnungsparameter
 - 5.2 Ergebnis der Analyse des Tonbandprotokollekorpus
 - 5.3 Validierung durch das Flugblattkorpus
 6. Fazit
- Literaturverzeichnis

1. Korpuslinguistik und Korpuspragmatik

1.1 Korpuspragmatik: Korpuslinguistik als Methode einer kultur- und sozialwissenschaftlich interessierten Linguistik

Die Korpuslinguistik ist eine zentrale Methode der Linguistik. Sie wird dazu benutzt, für fast alle Bereiche sprachwissenschaftlicher Forschung Daten zu konservieren, zu strukturieren und gezielt zu analysieren. War

die Korpuslinguistik zunächst vor allem eine zentrale Methode der Lexiko- und Grammatikographie, ist sie inzwischen auch aus dem Methodenrepertoire von Sozio- und Gesprächslinguistik, von Pragma- und Diskurslinguistik nicht mehr wegzudenken. Während die Korpuslinguistik in der systemorientierten Linguistik die Funktion hat, wiederkehrende Muster des Sprachgebrauchs zu identifizieren, die dann als Regularitäten oder Gebrauchsnormen gedeutet werden, werden in den kultur- und sozialwissenschaftlich interessierten Zweigen der Linguistik rekurrente sprachliche Muster mit kulturellen oder sozialen Phänomenen in Zusammenhang gebracht, je nach sprachtheoretischer Haltung werden sie entweder als deren Symptom oder als diese (mit-)konstituierend gedeutet. Wir wollen im Folgenden solche Ansätze als korpuspragmatisch bezeichnen.

Die Korpuspragmatik deutet signifikant häufig auftretende sprachliche Muster in Korpora als Ergebnis rekurrenter Sprachhandlungen der Autorinnen und Autoren der im Korpus enthaltenen Texte bzw. der sie autorisierenden Institutionen und Gruppen. Auf den ersten Blick erscheint es widersinnig, ausgerechnet mit dem Mittel der Korpuslinguistik pragmalinguistische Forschung betreiben zu wollen. Schließlich wird in der Linguistik die Pragmatik gerade als jene Disziplin beschrieben, die über das an der sprachlichen Oberfläche Zugängliche hinausgreift und in die Rekonstruktion von kommunikativem Sinn mit einbezieht. So bezeichnen zentrale Kategorien der traditionellen Pragmatik wie Illokution, Implikatur oder Präsupposition gerade solche linguistischen Konstrukte, die nicht an der sprachlichen Oberfläche identifizierbar sind – oder zumindest nicht vollständig. An diesem Punkt wird klar, dass die Korpuspragmatik etwas anderes ist als die Fortsetzung der traditionellen Pragmatik mit korpuslinguistischen Methoden. Sie bedient sich anderer Kategorien und anderer Strategien der (Re-)Konstruktion pragmatischer Bedeutungen. Sie ist neueren Einsichten der Pragmatik verpflichtet, nach denen sich pragmatische Informationen nicht nur auf der Ebene von Sprechakten oder aktuellen Texten manifestieren, sondern zeichenhaft manifest sind „im pragmatischen Mehrwert oder Gebrauchswert von Einheiten aller sprachlicher Strukturbereiche.“¹ Sie sucht nach pragmatischen Spuren an der (inzwischen rehabilitierten²) sprachlichen Oberfläche, nach Mustern, in die sich ein Gebrauchswert eingeschrieben hat.

Dabei ist die Korpuspragmatik methodologisch der quantitativen empirischen Sozialforschung verpflichtet. Sie ist darum bemüht, Kon-

1 FEILKE (2000: 78).

2 Vgl. den Sammelband FEILKE / LINKE (2009).

strukturen messbare Phänomene zuzuordnen. Anders als etwa in der Umfrageforschung sind diese Konstrukte keine Antwortmuster auf Itembatterien, sondern das Auftreten bestimmter sprachlicher Phänomene und ihre Verteilung. Der vorliegende Aufsatz befasst sich in methodologischer Hinsicht am Beispiel des Konstruktes *Stil* mit der Frage, auf welche Weise geeignete Kategorien zur Messung von Konstrukten identifiziert werden können. In eine korpuspragmatische Diktion übersetzt: Mit welchen Methoden ist es möglich, sprachliche Muster zu identifizieren, die in pragmatischer Hinsicht salient sind,³ sprachliche Muster also, die als das Ergebnis sprachlich-sozialer Handlungen gedeutet werden können.

1.2 Korpuslinguistik als Denkstil: corpus based vs. corpus driven

Zwar gilt die Korpuslinguistik als keine Subdisziplin der Linguistik sondern als eine Methode, korpuslinguistisches Arbeiten folgt aber einer eigenen Logik und generiert einen Denkstil, der viele Bereiche der Sprachwissenschaft nachhaltig verändert. Am ehesten der Vorstellung von Korpuslinguistik als einer Methode entspricht das *corpus-based*-Paradigma innerhalb der Korpuslinguistik. Digitale Korpora dienen demnach der Überprüfung von Forschungshypothesen. Die Hypothesen, die unabhängig von der Analyse des Korpus entwickelt wurden, formulieren Annahmen über interpretative Konstrukte, die mittels bereits bewährter interpretativer linguistischer Analysekatgeorien an einem Korpus überprüft werden sollen.

Diesem deduktiven Vorgehen steht die Möglichkeit eines induktiven Vorgehens zur Seite, das die Grundlage des *corpus-driven*-Paradigma bildet. Digitale Korpora sind hier nicht nur „Belegsammlungen oder Zettelkästen in elektronischer Form“, sondern ermöglichen eine eigene „korpuslinguistische Perspektive“.⁴ Statt eine Hypothese mit vorher festgelegten Analysekatgeorien zu überprüfen, werden in einem Korpus sämtliche Muster berechnet, die sich bei der Anwendung vorher festgelegter Algorithmen ergeben. Diese Muster werden im Anschluss kategorisiert. Damit geraten häufig Evidenzen in den Fokus, die entweder quer zu den vorher existierenden Erwartungen stehen und die Grundlage für neue Hypothesen sind, oder im besten Fall sogar solche

3 Denn natürlich verdanken sich nicht alle rekurrenten Muster in Textkorpora der pragmatischen Funktion der Texte. Dass am Anfang von Nominalgruppen signifikant häufig Artikel stehen, verdankt sich der Orientierung an den Regularitäten des Sprachsystems und dürfte in sozial- oder kulturwissenschaftlich interessierten Studien in den meisten Fällen nicht von Interesse sein.

4 PERKUHN/BELICA (2006: 2).

Evidenzen, die die Bildung neuer interpretativer linguistischer Analysekategorien nahelegen.⁵ Es ist dieses Potenzial datengeleiteter Analysen, das es erlaubt, der Korpuslinguistik über eine empirische Methode hinaus den Status eines Denkstils zuzuschreiben. Denn das Ernstnehmen empirischer Widerständigkeiten, die sich mit traditionellen linguistischen Kategorien nicht beschreiben lassen, birgt die Möglichkeit eines neuen Zugangs zu Sprache und den Kategorien ihrer Beschreibung.

Zwar verzichtet das datengeleitete Paradigma auf das Formulieren von Hypothesen und auf eine Festlegung auf bestimmte Analysekategorien, es ist jedoch offensichtlich, dass auch beim datengeleiteten Verfahren vorgängiges Wissen in den Forschungsprozess einfließt, und zwar:

1. durch die Wahl der Korpora,
2. hinsichtlich der Gestaltung der Algorithmen zur Musterberechnung,
3. bei der Festlegung dessen, was als linguistische Untersuchungseinheit (token) gelten soll, und
4. bei der Festlegung dessen, welche Einheitentypen eigentlich als potenzieller Bestandteil eines Musters aufgefasst werden sollen. Schließlich ist
5. auch das Kategorisieren der Daten im Anschluss an die Musterberechnung ein interpretativer Prozess, der zwar durch statistische Verfahren teilweise objektiviert werden kann; dennoch ist die Menge der Daten meist so umfangreich, dass eine weitere Reduzierung und Gewichtung im Sinne des Forschungsinteresses vorgenommen werden muss.⁶

Es ist für die Qualität einer datengeleiteten Analyse entscheidend, sich über das eingebrachte Vorwissen im Sinne der Forschungsfrage Rechenschaft abzulegen und nach möglichst objektiven Kriterien für jede der getroffenen Entscheidungen zu suchen.

1.3 Datengeleitete Korpusanalyse und Korpuspragmatik

Die von uns geleitete Forschergruppe *semtracks* bemüht sich im Rahmen eines drittmittelfinanzierten Forschungsprojekts mit dem Titel

-
- 5 Vgl. für eine ausführliche Diskussion von *corpus-based* und *corpus-driven*-Ansätzen BUBENHOFER 2009: 99ff.). Das *corpus-driven*-Paradigma war bereits bei SINCLAIR (1991) angedacht und wird bei TOGNINI-BONELLI (2001: 65ff.) explizit gemacht. Im deutschsprachigen Raum verfolgen z.B. Arbeiten von Kathrin Steyer (STEYER 2004, STEYER/LAUER 2007, STEYER/BRUNNER 2009) dieses Paradigma, die auf Konzepten und Methoden von Cyril Belica, Rainer Perkuhn, Holger Keibel und Marc Kupietz fußen (BELICA 1996, BELICA 2001, PERKUHN/BELICA 2006).
- 6 Vgl. MCENERY/XIAO/TONO (2006: 8–11).

„Bedeutungsspuren an der Oberfläche: Methoden der maschinellen Analyse semantischer Prägungen von Texten“ darum, das Potenzial datengeleiteter Methoden für die Korpuspragmatik auszuloten. Unsere Methode zur Identifizierung sprachlicher Muster mit pragmatischer Ladung ist die Methode des datengeleiteten Korpusvergleichs: Für in pragmatischer Hinsicht je kohärente Korpora werden jeweils datengeleitet Sprachgebrauchsmuster berechnet; im Anschluss wird anhand von Signifikanztests geprüft, welche der errechneten Muster typisch für die jeweiligen Korpora sind. Da die Korpora sich im Hinblick auf bestimmte außersprachliche Merkmale unterscheiden, können jene Muster, die für ein Korpus signifikant sind, als Ausdruck jener außersprachlichen Merkmale gedeutet werden, durch die sich das betreffende Korpus von den Vergleichskorpora unterscheidet.

Es ist offensichtlich, dass der Wahl der Korpora bei unserer Methode zentrale Bedeutung zukommt. Die Korpora sollten stets so gewählt sein, dass ausschließlich die durch die Forschungsfrage vorgegebene Variable variiert. So haben wir beispielsweise linguistische Indikatoren extremistischer Einstellungen anhand von Korpora berechnet, die im Hinblick auf außersprachliche Merkmale wie die Entstehungszeit der Texte (eine Legislaturperiode), die Textsorte (Pressemitteilung der Bundespartei), die intendierten Adressaten (Medienöffentlichkeit) und die Textintention (Stellungnahmen zu aktuellen politischen Fragen, Hinweis auf Veranstaltungen) übereinstimmen. Das einzige unterscheidende Merkmal war die Zugehörigkeit einer Partei zu einer Gliederungseinheit des politischen Spektrums.⁷

Da die Regularitäten einer Varietät (im Normalfall der Standardvarietät) in allen Korpora in gleicher Weise repräsentiert sind, sind all jene rekurrenten Muster in einem Korpus, die ausschließlich sprachsystematischen Charakter haben, für keines der Korpora signifikant. Solche Muster machen unserer Erfahrung nach zumeist mindestens 99,5% aller errechneten Muster aus. Die Systematisierung und Deutung der verbleibenden Muster bleibt dennoch ein aufwändiges und schwieriges Verfahren.

Im Folgenden wollen wir unsere Methoden am Beispiel einer Stilanalyse veranschaulichen. Hierfür ist es zunächst nötig, ausgehend von einer Konzeptspezifikation des Stilbegriffs zu einer korpuslinguistisch-operationalen Definition von *Stil* zu kommen. Darauf aufbauend werden wir die Algorithmen zur Musterberechnung entwickeln, geeignete Einheiten für die Musterbildung diskutieren und schließlich anhand

7 Zum zugrundeliegenden Extremismusbegriff und den Ergebnissen vgl. EBLING et al. 2011.

zweier Korpora zur 68er-Bewegung erproben. Das der exemplarischen Stilanalyse zugrundeliegende Erkenntnisinteresse liegt darin zu untersuchen, inwieweit lebensstilistische Unterschiede in der 68er-Bewegung kommunikationsstilistische Korrelate haben.

2. Zum Stilbegriff

Stil ist ein Begriff im Schnittpunkt zahlreicher Wissenschaften: in Kunstwissenschaft, Literaturwissenschaft, Soziologie, Psychologie und weiteren Disziplinen, darunter auch der Linguistik, ist der Stilbegriff ein analytisches Konzept, mit dem die spezifische Art und Weise der Durchführung von Handlungen oder der Ergebnisse von Handlungen in den Blick genommen werden. Viele Definitionen stimmen darin überein, dass sie Stil als eine (Aus-)Wahl bestimmter Elemente aus einer Menge quasi-synonymer Ausdrucksmöglichkeiten aus dem gesamtgesellschaftlichen Zeichenrepertoire bestimmen. Diese (Aus-)Wahl hat soziale Signifikanz insofern, als sie die Identitäten der Träger von Stilen mitkonstituiert, Interpretationsrahmen transportiert und handlungsleitende Kontexte generiert.⁸ Stil ist ein holistisches Konzept, weil es den Gebrauch von Zeichen unterschiedlicher Materialität, Medialität und Komplexität als Ausdruck eines allen einzelnen Wahlen zugrundeliegenden gemeinsamen ästhetischen Schemas deutet und so gewissermaßen eine tiefensemantische Homologie zwischen den unterschiedlichen Zeichen behauptet.

Wir sind in dieser exemplarischen Studie daran interessiert, in Korpora messbare Stile in Relation zur außersprachlichen Variable *Lebensstil* zu setzen. Wir werden uns daher im Folgenden auf einige kurze Ausführungen zum soziologischen Lebensstilkonzept und zum Stilkonzept der interaktionalen und Textlinguistik beschränken.

2.1 Das Stilkonzept der Lebensstilforschung

Das Lebensstilkonzept ist seit den 1980er Jahren zum zentralen Konzept nicht nur der deutschsprachigen Ungleichheitsforschung in der Soziologie avanciert und hat Konzepte wie *Klasse* oder *Schicht* weitgehend verdrängt.⁹ Diese Entwicklung korrespondiert einem Wandel in der Selbstdefinition von in modernen Gesellschaften lebenden Menschen. Sie bestimmen ihre eigene soziale Lage verstärkt über ihre Lebensweise und immer weniger über die Selbst- oder Fremdzuschrei-

8 Vgl. die unterschiedlichen Definitionen in HEINEMANN (2009: 1612).

9 Vgl. NASSEHI (1998: 56f.).

bung zu einer sozialen Schicht.¹⁰ Entscheidende Anregungen für diesen Paradigmenwechsel erhielt die Ungleichheitsforschung dabei aus der französischen Kulturosoziologie Pierre Bourdieus. Ihm zufolge erwerben die Mitglieder einer Gesellschaft im Laufe ihrer Sozialisation ein System dauerhafter Dispositionen, die als „Erzeugungs- und Ordnungsgrundlagen für Praktiken und Vorstellungen“ fungieren.¹¹ Sie steuern, welche Verhaltensweisen einem Menschen als angemessen, welche Meinungen als akzeptabel, welche symbolischen Formen als geschmackvoll und welche Wissensbestände als erstrebenswert gelten. Dieses evaluative System nennt Bourdieu den Habitus. Geleitet von diesem wähle der Mensch aus der Reihe ihm in seiner Kultur zur Verfügung stehenden Handlungsmöglichkeiten aus und „entscheidet“ sich so für einen bestimmten Lebensstil. Dabei sind in Bourdieus Theorie die objektiven ökonomischen Möglichkeiten der entscheidende Faktor, der die dispositive Prägung während der Sozialisation beeinflusst:

Die charakteristischen Strukturen einer bestimmten Klasse von Daseinsbedingungen sind es nämlich, die über die ökonomische und soziale Notwendigkeit [...] die Strukturen des Habitus erzeugen, welche wiederum zur Grundlage der Wahrnehmung und Beurteilung aller späteren Erfahrung werden.¹²

Auch wenn in der deutschsprachigen Lebensstilforschung die objektiven ökonomischen Bedingungen eine weitaus geringere Rolle spielen, sind Bourdieus Überlegungen dennoch auch hier grundlegend für eine kulturosoziologische Analyse sozialer Milieus.

Auch für die Analyse sozialer Bewegungen sind Habitus und Lebensstil wichtige Kategorien, wenn es darum geht, die Verfestigung von Bewegungsidentitäten zu erklären. Schon Bourdieu hatte festgestellt, dass bei der Mobilisierung sozialer Gruppen auch eine habituell motivierte Abstimmung von Praktiken in Gang kommen müsse. Die „Unterfangen kollektiver Mobilisierung“ blieben erfolglos, „wenn nicht ein Mindestmaß an Übereinstimmung zwischen den Mobilisierenden (Propheten, Rädelsführern usw.) und den Dispositionen derjenigen gegeben ist, die sich in deren Praktiken und Äußerungen wiedererkennen [...]“.¹³ Die hervorragende Bedeutung visueller Codes für die Formierung einer kollektiven Identität der 68er-Bewegung hat Kathrin Fahlenbrach in ihrer Studie „Protest-Inszenierungen“ herausgearbeitet.¹⁴ In einem gleichnamigen Aufsatz resümiert sie:

10 Vgl. HRADIL (2001: 46).

11 BOURDIEU (1993: 98).

12 BOURDIEU (1993: 101).

13 BOURDIEU (1993: 111).

14 Vgl. FAHLENBRACH (2002).

Wie in keiner anderen sozialen Bewegung zuvor erhalten in dieser kulturhistorischen Konfliktsituation habituelle Repräsentationsformen einen zentralen Stellenwert für die Entwicklung einer kollektiven (Protest-)Identität: neben visuellen Symbolen und Emblemen, die ähnlich wie in traditionellen Bewegungen genutzt werden (Fahnen, Plakate mit Parolen usw.), umfasst dies alle Formen der kollektiven und personalen Selbstdarstellung: Bilder, Kleidung, Frisuren, Gestik,¹⁵ Mimik, usw., sowie neue symbolische Formen öffentlicher Protestaktionen.

Im Rahmen unserer exemplarischen Studie interessiert uns die Frage nach dem Zusammenhang von milieuspezifischen Lebensstilen innerhalb der 68er-Bewegung und den Kommunikationsstilen in milieuspezifisch definierten Korpora.

2.2 Stilkonzepte in der Linguistik

Den Stilkonzepten der Linguistik ist gemeinsam, dass sie Stil als eine Dimension von Äußerungen ansieht, die die Interpretation seitens des Rezipienten steuert, ohne zur sprachlichen Primärstruktur der Äußerung zu gehören. Stil ist damit eine Art pragmatische Zusatzinformation.¹⁶ Diese Bestimmung trifft auf die Textlinguistik gleichermaßen zu wie auf die interaktionale Linguistik. In Letzterer wurde der Terminus der Sozialstilistik der Kommunikation geprägt. Im Gegensatz zum Konzept des Lebensstils, dem eine dauerhafte Disposition zur Wahl bestimmter expressiver Formen zugrunde liegt und der auch mittels nicht-transitorischer Zeichen konstituiert wird, werden kommunikative soziale Stile aus dem kommunikativen Handeln heraus konstruiert. Der Terminus Stil bezeichnet hier die interaktive und daher sozial bedeutsame Art und Weise der Handlungsdurchführung. Stile sind Zeichenensembles mit indexikalischem Wert, denn sie signalisieren, wie das Handeln zu interpretieren ist. Sie verweisen auf Interpretationsrahmen, die bei der Kategorisierung und Typisierung von Handlung, Kontext und Interaktionspartnern zur Anwendung kommen.¹⁷ Sprecher verfügen über ein mehr oder weniger großes Repertoire kommunikativer sozialer Stile, aus dem sie je nach Situation oder kommunikativem Bedürfnis wählen können. In Sprechstilen sind Phänomene aus so unterschiedlichen linguistischen Subsystemen wie Rhetorik, Lexiko-Semantik, Syntax, Phonetik-Phonologie und Prosodie gebündelt. Dennoch werden sie von Rezipienten holistisch, d.h. ganzheitlich wahrgenommen und funktional

15 FAHLENBRACH (2007: 12).

16 Vgl. hierzu FIX (1987: 133). FIX (1996: 310) definiert (SANDIG 1995: 28 zuspitzend): „Funktion von Stil wäre also, kurzgefaßt, ein ‚Was‘ durch ein ‚Wie‘ im Hinblick auf ein ‚Wozu‘ auszudrücken“. Vgl. zudem HEINEMANN (2009: 1616).

17 Vgl. SELTING (1997: 12).

interpretiert. D. h. das sprachliche Merkmalsbündel wird intuitiv als Index für einen Interpretationsrahmen aufgefasst.¹⁸

2.3 Korpuslinguistische Operationalisierung des Stilbegriffs

Schon aus theoretischen Erwägungen heraus ist der Text die grundlegende Analyseeinheit der Stilanalyse. Fix erklärt die Wechselbeziehung von Stil und Text wie folgt:

Die Betrachtung von Musterhaftigkeit im Stil ist von der Musterhaftigkeit der Texte nicht zu trennen. Dies bedeutet einerseits, dass es Stil nur im Textzusammenhang gibt und sprachliche Mittel außerhalb des Textes stilistisch nicht eingeordnet werden können, und andererseits, dass die reale Existenz eines Textemplars auch vom Vorhandensein eines einheitlichen Stils abhängt. Ohne diesen sind Textmusterbezogenheit und Textualität eines Textes nicht zu erkennen.¹⁹

Wenn sich Stil in der Selektion einer Variante von paradigmatischen Alternativen ausdrückt²⁰, so können durch Stilvergleiche diese Alternativen sichtbar gemacht werden. So schlägt Fix auch das Verfahren des Vergleichs vor, um Elemente eines Stils in Texten zu erfassen.²¹ Denn die Kombination von „Wie“ und „Wozu“ zeige sich im Unikalen des Textes erst im Vergleich mit dem Überindividuellen. Jeder Text ist unikal, verfügt jedoch über einen Stil, der sich an überindividuellen Erfahrungen orientiert. Durch den Vergleich des Textes mit einem „individuellen Konstrukt“²², das sich aus der Lektüreerfahrung ergibt²³, kann das Überindividuelle des Textes erfasst werden.

In automatisierten Verfahren wird das individuelle Konstrukt, das sich aus der Lektüreerfahrung ergibt und das als Maß für die Abweichung bzw. Kohärenz eines Textes mit seinen Co-Texten dient, durch die Berechnung der Verteilung aller in einem Korpus enthaltenen Muster ersetzt. Die Lektüreerfahrung wird damit also durch ein statistisches Maß objektiviert. Unsere operationale Definition von Stil lautet daher: Stil ist in korpuslinguistischer Perspektive eine Menge sprachlicher Muster, durch die sich eine Menge an Texten durch eine andere Menge von Texten signifikant unterscheidet.

18 Vgl. SELTING (1997: 14).

19 FIX (2009: 1312), vgl. auch FIX (2005).

20 HOCKET (1958: 556) formuliert diesen Gedanken anschaulich wie folgt: „Roughly speaking, two utterances in the same language which convey approximately the same information, but which are different in their linguistic structure, can be said to differ in style: ‚He came soon‘ and ‚He arrived prematurely‘.“

21 FIX (1991).

22 FIX (1991: 142).

23 FIX (1991: 145).

Das von Fix beschriebene Verfahren der intuitiven Bestimmung von Stil ist in datengeleiteter Perspektive einer sehr komplexen Clusteranalyse vergleichbar, in der alle in einem Korpus vorkommenden Muster Variablen sind, die potentiell zur Gruppierung der Texte des Korpus in mehr oder weniger diskrete Gruppen beitragen. Das im Rahmen dieses Aufsatzes vorgestellt Vorgehen ist im Vergleich dazu einfacher: Wir berechnen ausgehend von der Annahme der Existenz zweier unterschiedlicher Stile anhand zweier Korpora, welche Muster für die jeweiligen Korpora signifikant sind. Anders als in einer Clusteranalyse geben wir also die Gruppierung der Texte vor, beziehen aber ebenso wie in der Clusteranalyse sämtliche Muster beider Korpora als Variablen in die Analyse ein.

3. Die *semtracks*-Stilanalyse

Es gibt eine Tradition quantitativer Verfahren der Stilanalyse, bei der formale Aspekte wie Silbenzahl pro Wort, Wort- und Satzlängen, Wortfolgen, Satzgliedtypen, Wortklassen und generell lexikalisches Material berücksichtigt werden.²⁴ Auch auf morphologischer Ebene, z.B. was die Wortbildung betrifft, finden sich messbare formale Aspekte.²⁵ Grundlage sind Korpusvergleiche zwischen dem stilistisch zu untersuchenden Text und einem repräsentativen Vergleichskorpus, um Frequenzen der genannten formalen Phänomene miteinander vergleichen zu können.²⁶

Solche quantitative Analysen formaler Besonderheiten eines Textes haben den Nachteil, viele stilbestimmende Ebenen zu ignorieren und vor allem das Zusammenspiel mehrerer Faktoren zu wenig zu berücksichtigen. Dennoch knüpfen wir an die Idee an, messbare linguistische Elemente als grundlegende Analyseeinheiten zu wählen und mit statistischen Verfahren zu eruieren, welche Kombinationen von Einheiten für einen Text oder ein Korpus in Kontrast zu Vergleichskorpora typisch sind.

Wir halten es dabei nicht für sinnvoll, einen Purismus der Oberfläche zu praktizieren. Die *semtracks*-Stilanalyse bezieht Oberflächenphänomene auf der Ebene der Wortformen und Satzzeichen ebenso mit ein, wie weitergehend interpretative linguistische Kategorien. Zudem erscheint uns eine Beschränkung auf Unigramme, das heißt auf Muster, die nur aus einem Element bestehen, nicht sinnvoll, da bei Unigrammanalysen die Kontextspezifik der Verwendung von Lexemen nicht be-

24 SPILLNER (2009: 1758f.).

25 Vgl. HANDLER (2009).

26 Vgl. dazu auch Arbeiten im Rahmen der *Corpus Stylistics*, z.B. von SEMINO/SHORT (2004), HOOVER (2007).

rücksichtigt wird. Die Gefahr, dass der pragmatische Wert eines Wortgebrauchs nicht erfasst wird, ist bei diesem Verfahren besonders groß. Unser Vorgehen lässt sich vielmehr als *komplexe n-Gramm-Analyse* beschreiben.

3.1 Komplexe n-Gramm-Analyse

N-Gramme sind Einheiten, die aus n Elementen bestehen. Normalerweise werden n-Gramme als Folge von Wortformen verstanden. Im Rahmen einer n-Gramm-Analyse werden alle im Korpus vorkommenden n-Gramme berechnet, wobei bestimmte Parameter wie Länge der Mehrworteinheit (aus zwei, drei oder mehr Wörtern bestehend) oder Spannweite (sind Lücken zwischen den Wörtern erlaubt?) festgelegt werden.²⁷ Die von unserer Forschergruppe entwickelte n-Gramm-Analyse betrachtet jedoch nicht nur Wortformen als Einheiten, sondern auch weitere interpretative linguistische Kategorien. Dies können zum einen Elemente sein, die sich auf die Tokenebene beziehen und die Wortform funktional oder semantisch deuten (als Repräsentant einer Wortart oder als Teil einer semantischen Klasse). Zum anderen aber auch Elemente, die über die Tokenebene hinausgreifen, etwa das Tempus oder die Modalität einer Äußerung (direkte vs. indirekte Rede).

Welche Elemente in die Analyse mit einbezogen werden, hängt einerseits von der jeweiligen Forschungsfrage ab, andererseits forschungspraktisch auch davon, welche Ressourcen für die Annotation des Korpus zur Verfügung stehen. Bei standardsprachlichen Korpora können Lemma- und Wortarteninformationen durch Tagger wie den *TreeTagger* leicht und effizient annotiert werden. Eine Wortformenfolge wie „Ich glaube, dass“ hat dann in einem XML-annotierten Korpus etwa folgende Form:

```
<w pos="PPER" lemma="ich">Ich</w>
<w pos="VVFİN" lemma="glauben">glaube</w>
<w pos="$. " lemma=",">,</w>
<w pos="KOUS" lemma="dass">dass</w>
```

Berechnet man nun beispielsweise Tetragramme, die nicht nur die Wortformen, sondern auch Lemmata und Wortarteninformationen als weitere Elemente mit einzubeziehen, dann ergeben sich bei drei Dimensionen $3^4=81$ Vier-Einheiten-Kombinationsmöglichkeiten:

Ich glaube , dass
ICH GLAUBEN , DASS

²⁷ Vgl. BUBENHOFER (2009: 149ff.).

PPER glaube , dass
 PPER GLAUBEN, dass
 Ich VVFIN , dass
 Ich glaube , KOUS
 PPER VVFIN , dass
 ...

Jedes der Tetragramme, das sich in einem der beiden Korpora findet, kann nun als eine Variable aufgefasst werden, durch die sich die beiden Korpora potenziell stilistisch unterscheiden. Es ist dann eine stilistisch relevante Variable, wenn sie in einem der beiden Korpora signifikant häufiger vorkommt als im Vergleichskorpus. Zur Berechnung der Signifikanz des Frequenzunterschieds benutzen wir den Log-Likelihood-Test.²⁸

Durch die freie Verfügbarkeit von Taggern hat die eben skizzierte Berechnungsmethode das Potenzial zu einer Standardmethode der automatisierten Stilanalyse zu werden. Reichere Ergebnisse sind freilich zu erwarten, wenn zusätzliche Elemente in die n-Gramm-Analyse einbezogen werden. Diese können einerseits wissensbasiert generiert werden, z.B. die Valenz von Verben anhand von Valenzlexika oder die Zugehörigkeit eines Lexems zu einem Wortfeld anhand der Bestimmungen von Dornseiff,²⁹ andererseits können aber auch anhand datengeleiteter Analysen selbständig relevante Kategorien aufgespürt und definiert werden. Insbesondere auf der Tokenebene ist dieses Verfahren in effizienter Weise durchführbar.

3.2 Datengeleitete Findung relevanter Kategorien auf der Tokenebene

Der einfachste Zugang zu linguistischen Spezifika der Teilkorpora auf der Tokenebene ist die Berechnung von typischen Lexemen. Es werden dabei für zwei Teilkorpora jeweils Frequenzlisten aller vorkommenden Lexeme berechnet. Um eine leichtere Vergleichbarkeit funktional ähnlicher Lexeme zu ermöglichen, ist es in vielen Fällen sinnvoll, Listen gesondert nach den Wortarteninformationen zu berechnen. Anschließend werden pro Lexem die Frequenzen in den beiden Teilkorpora verglichen und die Signifikanz des Frequenzunterschieds berechnet. Das Resultat sind nach Signifikanzwert geordnete Listen von Lexemen,

28 MANNING/SCHÜTZE (2002: 174).

29 DORNSEIFF (2004).

die jeweils für das eine Teilkorpus im Vergleich zum anderen Teilkorpus typisch sind.

In einem weiteren Schritt können die typischsten Lexeme in sog. *Wortwolken* visualisiert werden, wobei die Typik des Lexems durch die Größe des abgebildeten Lexems dargestellt wird. Dies erlaubt einen raschen Überblick über die lexikalische Spezifik der Teilkorpora. Dieses Hilfsmittel kann dazu eingesetzt werden, weitere semantische oder funktionale Wortklassen zu identifizieren, die im Anschluss im Korpus annotiert werden.³⁰ Auf diese Weise können die Klassenzugehörigkeit eines Tokens bzw. die jeweilige Dimension der Klasse, die das Token repräsentiert, zum Element von n-Grammen und somit zu einem Bestandteil des Stils werden.

Diese recht abstrakte Darstellung unserer Methode der Stilanalyse soll im Folgenden anhand einer exemplarischen Analyse zu unterschiedlichen kommunikativen sozialen Stilen während der 68er-Bewegung illustriert werden.

4. Das GerMov-Korpus

Die Untersuchungen werden anhand des *GerMov-Korpus*, einem Korpus zur gesprochenen und geschriebenen Sprache der 68er-Bewegung durchgeführt. Das Korpus wurde im Rahmen einer umfangreichen Studie zum Einfluss von 68er-Bewegung und Alternativmilieu auf die Kommunikationsgeschichte der Bundesrepublik Deutschland erstellt.³¹ Bei der Erstellung des Korpus und seiner Subkorpora waren zunächst außersprachliche Gesichtspunkte, in einem zweiten Schritt textlinguistische Überlegungen leitend.

4.1 Kriterien bei der Korpuserstellung

Das Korpus sollte es u. a. erlauben, unterschiedliche Stile der verbalen face-to-face-Interaktion innerhalb der 68er-Bewegung zu rekonstruieren. Dabei wurde ausgehend von der Forschung zum Kleidungsverhalten³² und zur medialen Vermittlung expressiver Formen des Protests³³ von einer lebensstilistischen Dualität innerhalb der Bewegung ausgegangen, die ihre Wurzeln auch in konkurrierenden Ideologien hatte.³⁴

30 Vgl. für Beispiele BUBENHOFER et al. (2009) und BUBENHOFER/KLIMKE/SCHARLOTH (2009) und BUBENHOFER/SCHARLOTH (2010).

31 Vgl. SCHARLOTH (2011).

32 Vgl. GROB (1985).

33 Vgl. FAHLENBRACH (2002).

34 Vgl. zu der folgenden groben Skizze ausführlicher Kapitel 3 in SCHARLOTH (2011). Die Bezeichnung der Stile und Milieus folgt GROB (1985) und SCHARLOTH (2007).

Auf der einen Seite standen die Träger eines intellektuell-avantgardistischen Stils. Bei ihnen handelte es sich um Angehörige unterschiedlicher sozialer Gruppen, die während der 68er-Bewegung aber intensiv kooperierten: zum einen die Studierenden, vornehmlich solche, die in linken Studentenverbänden organisiert waren, zum anderen Linksintellektuelle, die in Politik, Universität, Verwaltung oder im kulturellen Sektor bereits Karriere gemacht hatten, die sich beispielsweise in Republikanischen Clubs zusammenfanden. Sie pflegten einen auf symbolische Distinktion zunächst weitgehend verzichtenden Lebensstil, trugen Anzug oder Freizeitkleidung (Hemd und Pullovern, Jacket und Cordhose) und praktizierten Lebensformen wie andere Bürger ihrer Berufsgruppen. Nur in einem Bereich legten sie Wert auf Unterscheidung: Sie inszenierten sich als intellektuelle Informations- und Diskussionselite. Auf der anderen Seite standen die Träger eines hedonistischen Selbstverwirklichungsstils, der in Kommunen und Subkulturen geprägt wurde. Sie entdeckten den eigenen Körper als zentrales Medium des expressiven Protestes, griffen – ähnlich den amerikanischen Hippies – tief in den Fundus von Kostümverleihen und Second-Hand-Läden, spielten mit Nacktheit und Schmuck, ließen sich Bärte und Haare wachsen und praktizierten eine ostentativ informelle Körpersprache. Sie verschmolzen antibürgerliche symbolische Formen mit denen jugendlicher Populärkultur zu einem sich als individualistisch verstehenden, lustbetonten Lebensstil: Die Revolution sollte bei jedem Einzelnen beginnen und vor allem Spaß machen. Während die intellektuellen Avantgarden das Ziel der 68er-Bewegung in einer Umwälzung der Besitz- und Produktionsverhältnisse sahen, begriffen die hedonistischen Kommunarden also die Bewegung als Chance für die Entwicklung und Praktizierung neuer Lebensformen, die eine gesellschaftliche Veränderung zwangsläufig mit sich bringen würde. Die Träger dieser Lebensstile formierten im Verlaufe der 68er-Bewegung bis weit in die 70er Jahre hinein bestehende soziale Milieus.

Um zu überprüfen, ob den unterschiedlichen Lebensstilen der Angehörigen der 68er-Bewegung auch unterschiedliche Kommunikationsstile korrespondieren, war es also nötig, Texte nach dem Kriterium der *Milieuzugehörigkeit* ihrer Produzentinnen und Produzenten zu sammeln. Neben dem Kriterium der Milieuzugehörigkeit war das der *Medialität* bzw. das davon abgeleitete Kriterium der *Textsorte* von zentraler Bedeutung. Da anhand des Korpus Stile der verbalen face-to-face-Interaktion rekonstruiert werden sollten, wurde gezielt nach solchen Texten gesucht, die als Repräsentationen gesprochener Sprache gelten können. Um ein möglichst hohes Maß an Homogenität zu erreichen, wurde zudem das Kriterium der *Kommunikationssituation* be-

rücksichtigt. Da davon ausgegangen werden kann, dass es in stilistisch heterogenen Gruppen trotz Distinktionsbedürfnis zu stilistischen Akkomodationen in der Kommunikation kommt, sollten nur solche Kommunikationssituationen berücksichtigt werden, in denen Kommunikationspartner des gleichen Milieus miteinander kommunizieren.

Die Kriterien der Milieuzugehörigkeit der Textproduzenten, der Medialität / Textsorte und der Kommunikationssituation setzte der Textauswahl sehr enge Grenzen. Die einzige Textsorte, für die hinsichtlich aller Kriterien eine hinreichende Menge an Texten gefunden werden konnte, waren Tonbandprotokolle. Insgesamt konnten 29 Tonbandprotokolle aus den Jahren 1967 bis 1969 in Archiven und zeitgenössischen Buch- und Zeitschriftenpublikationen gefunden werden, davon stammen 21 aus dem hedonistischen Selbstverwirklichungsmilieu, 8 aus dem linksintellektuellen Milieu. Die Zuordnung erfolgte beim Linksintellektuellenmilieu anhand der identifizierbaren Gesprächsteilnehmer und deren Zugehörigkeit zu politischen Gruppen, die den Milieu eindeutig zuzuordnen waren. Die Protokolle aus dem Kommunemilieu waren ausnahmslos als solche betitelt und wurden in szenetypischen Kontexten publiziert, was auch hier eine zweifelsfreie Zuordnung ermöglichte. Leider standen die Tonbandaufzeichnungen der in das Korpus aufgenommenen Gespräche nicht mehr zur Verfügung, sondern lediglich ihre Abschriften. Da die Transkriptionen von unterschiedlichen Händen stammen, kann auch nicht davon ausgegangen werden, dass sie einheitlich und nach gleichen Kriterien erfolgte. So variiert beispielsweise der Umgang mit überlappenden Turns oder nichtstandardsprachlichen Ausdrucksmitteln stark. Bei einigen Protokollen wurden auch die Sprecherinnen und Sprecher der einzelnen Turns nicht identifiziert, was interaktionale Analysen erheblich erschwert. Aus diesen Gründen ist das GerMov-Tonbandprotokollekorpus nicht mit Korpora zur gesprochenen Sprache des gegenwärtigen Deutsch vergleichbar.

Mit dem GerMov-Korpus sollten neben der Frage nach der kommunikationsstilistischen Vielfalt weitere offene Fragen zur Kommunikationsgeschichte beantwortet werden. Ein zentrales Anliegen war beispielsweise die Untersuchung der Hypothese, nach der die 68er-Bewegung gescheitert sei, weil ihre wichtigsten Gruppen sich eines zu komplizierten Kommunikationsstils bedient hätten. Sprachgebrauch sollte also in Relation zu dem außersprachlichen Kriterium des Verhältnisses von Äußerndem und Adressaten gesetzt werden. Zudem sollte untersucht werden, ob sich Kommunikationsstile im Verlauf der 68er-Bewegung veränderten. Für das zweite Subkorpus waren also die *Gruppenspezifika* der Texte, ihre *Adressaten* und ihre *Entstehungszeit*

zentrale außersprachliche Kriterien bei der Korpuserstellung. Für die Beantwortung der genannten Forschungsfragen schien es zudem sinnvoll, ein im Hinblick auf die *Textsorte* möglichst homogenes Teilkorpus zu erstellen, um ein möglichst hohes Maß an Vergleichbarkeit zu gewährleisten. Denn die Unterschiede in autoren-, adressaten- oder zeitspezifischen Teilkorpora dieses Subkorpus können nur dann als Unterschiede im Sprachgebrauch unterschiedlicher Autoren, gegenüber unterschiedlichen Adressatengruppen oder als typisch für eine bestimmte Phase der 68er-Bewegung gelten, wenn andere Variablen wie etwa die Textsorte als Ursache für die Differenzen im Sprachgebrauch ausscheiden. Eine Textsorte der Massenkommunikation, die eine zu- meist eindeutige Adressatenzuordnung nach den für die 68er-Bewegung grundlegenden Adressatengruppen (Angehörige der Bewegung, Studierende allgemein, Bevölkerung, Medienöffentlichkeit) erlaubt, ist das Flugblatt. Die Beschränkung auf Flugblätter ließ sich zudem dadurch rechtfertigen, dass Flugblätter in der 68er-Bewegung ein zentrales Medium zur Herstellung von Gegenöffentlichkeit waren und daher davon ausgegangen werden kann, dass sich in ihnen auch die Besonderheiten des Sprachgebrauchs verdichtet haben.

4.2 Das GerMov-Korpus und seine Subkorpora

Ausgehend von den skizzierten Fragestellungen und orientiert an den genannten Kriterien der Textauswahl gliedert sich das GerMov-Korpus in zwei Subkorpora: ein Subkorpus, das ausschließlich Tonbandprotokolle der 68er-Bewegung enthält, und ein Subkorpus mit Flugblättern. Das Tonbandprotokollkorpus enthält 29 Protokolle mit 59.879 laufenden Wortformen. Tabelle 1 zeigt die Milieuspezifische Verteilung von Texten und Frequenzen.

Das Flugblattkorpus enthält 508 Texte mit 213.010 laufenden Wortformen. Tabelle 2 zeigt die Verteilung der Texte auf die verschiedenen Positionen des politischen Spektrums.

<i>Milieu</i>	<i>Anzahl Texte</i>	<i>Anzahl Wortformen</i>
Hedonistisches Selbstverwirklichungsmilieu	21	27.736
Linksintellektuelles Milieu	8	32.143

Tab. 1: Übersicht über das GerMov-Tonbandprotokollekorpus

<i>Politisches Spektrum</i>	<i>Anzahl Texte</i>	<i>Anzahl Wortformen</i>
Radikale Linke	272	128.537
Alte Linke	1	318
Spirituelle Linke	2	646
Gemäßigte Linke	157	56.543
Linksliberal	9	4.297
Liberal	16	7.411
Linkskritisch	25	7.604
Konservativ	15	4.983
Rechtskonservativ	4	1.240
Rechtsextrem	2	421
kann nicht zugeordnet werden	4	1.010

Tab. 2: Übersicht über das GerMov-Flugblattkorpus

Für die vorliegende Studie wird das Flugblattkorpus dazu herangezogen, um die Ergebnisse der Analyse des Tonbandprotokollekorpus zu validieren. Hierfür ist eine Filterung der Texte nötig: Da das Tonbandprotokollekorpus ausschließlich Gespräche der radikalen Linken enthält, muss das Flugblattkorpus entsprechend gefiltert werden. Tabelle 3 zeigt, zu welchen Gruppierungen der radikalen Linken das Flugblattkorpus Texte enthält.

<i>Gruppierung</i>	<i>Anzahl Texte</i>	<i>Anzahl Wörter</i>
3. Welt - Gruppe	1	158
Aktionsrat zur Befreiung der Frauen	4	1.542
Anti-Notstandskampagne für Bundeswehrangehörige	1	214
EXTRA-Dienst	1	609
Freunde der Publizistik	3	1.087
Kuratorium Notstand der Demokratie	2	597
Lehrlinge	3	1.667
nicht in einer Gruppe organisierte Aktivisten	38	26.712
Republikanische Hilfe	1	287
Republikanischer Club	2	3.605
Schüler	4	1.140
SDS	49	2.1487
spontaneistisch-antiautoritäre Studentengruppe	58	32.711
Studentenvertretung (AStAs, Fachschaften etc.)	43	16.369
studentische Arbeitsgruppe	9	2.400
Vietnam-Solidaritätskomitee	1	263
Zusammenschluss mehrerer APO-Gruppierungen	8	3.676
Kommune I	23	6.316
Subkultur (Haschrebellen, weitere Kommunen, Blues)	19	7.387

Tab. 3: Gliederung des Flugblatt-Subkorpus „radikale Linke“ nach politischen Gruppierungen

Um eine möglichst hohes Maß an Vergleichbarkeit zu gewährleisten wurden in das Untersuchungskorpus nur Flugblätter jener Gruppierungen aufgenommen, die eindeutig einem der beiden im Tonbandprotokollekorpus repräsentierten Milieus zugeordnet werden können. Für den intellektuell-avantgardistischen Stil sind dies die Flugblätter des SDS, der spontaneistisch-antiautoritäre Studentengruppen und des Republikanischen Clubs, insgesamt 109 Flugblätter mit 57.803 laufenden Wortformen. Das Teilkorpus von Flugblättern aus Kommunen und Subkultur ist dagegen mit 42 Texten und 13.703 Wortformen deutlich kleiner.

	<i>Tonbandprotokolle</i>	<i>Flugblätter</i>
Hedonistisches Selbstverwirklichungsmilieu	27.736 (21 Protokolle)	13.703 (42 Texte)
Linksintellektuelles Milieu	32.143 (8 Protokolle)	57.803 (109 Texte)

Tab. 4: Übersicht über die Größe der Teilkorpora

4.3 Annotation

Das GerMov-Korpus wurde wie oben beschrieben mit Hilfe des TreeTaggers³⁵ tokenisiert, mit Wortarten-Informationen annotiert und lemmatisiert. Beim verwendeten Tagset handelt es sich um das Stuttgart-Tübingen-Tagset (STTS).³⁶ Darüber hinaus wurden die folgenden Kategorien auf der Token-Ebene annotiert:

- a. *Kommunikationsverben*: Da es sich beim Tonbandprotokollekorpus um ein Korpus der gesprochenen Sprache handelt, kann angenommen werden, dass Frequenz und Art des Gebrauchs von Kommunikationsverben stilrelevant sind. Kommunikationsverben wurden gemäß der von HARRAS et al. (2004) vorgeschlagenen Kategorisierung als KOMMVERB annotiert.
- b. *Intensivierer / Gradpartikel*: Intensivierer kodieren Emotionen und den Grad von Überzeugungen, bzw. der Rigorosität, mit der sie vertreten werden. Sie können in unterschiedliche Klassen eingeteilt werden.³⁷ Als stilrelevant erscheinen uns einer-

³⁵ SCHMID (1994).

³⁶ SCHILLER et al. (1995).

³⁷ Vgl. OS (1989), BIEDERMANN (1969), BIERWISCH (1987).

seits die Intensivierer des absoluten (z.B. *absolut, gänzlich, grundlegend, gründlich*) und des extrem hohen Intensivierungsbereichs (z.B. *höchst, äußerst, zutiefst*), die wir als GRAD_HOCH im Korpus annotiert haben; andererseits die des gemäßigten (z.B. *ziemlich, recht, einigermaßen*), des abschwächenden (z.B. *verhältnismäßig, relativ, eher*) und des minimalen (z.B. *wenig, kaum, schwerlich*) Intensivierungsbereichs, die wir als GRAD_RELATIV annotiert haben.

Schlagwörter der neuen Linken: Vergleicht man das Tonbandprotokollkorpus mit einem Gegenwartskorpus politischer Diskussionen aus einem Onlineforum³⁸ im Hinblick auf das Vorkommen von Nomen, so zeigt sich (vgl. Abbildung 1), dass im GerMov-Korpus vergleichsweise viele Nomen verwendet werden, die aus dem Bereich Marxismus / Neue Linke kommen. Es kann daher davon ausgegangen werden, dass Frequenz und Art und Weise des Gebrauchs von Schlagwörtern der Neuen Linken stilrelevant sind. Zur genaueren Bestimmung dessen, was als Schlagwörter der Neuen Linken gelten kann, wurden die Lemmata aus zeitgenössischen Wörterbüchern zur Sprache der Neuen Linken³⁹ ausgewertet und als SCHLAG annotiert. Innerhalb dieser Gruppe von Wörtern wurden solche, die sich als Kampfvokabular im Sinne eines sich als revolutionär verstehenden Klassenkampfes klassifizieren ließen, zusätzlich als KAMPF annotiert.

38 Es handelt sich um ein Korpus, das aus sämtlichen thematischen Strängen des Schweizer Online-Diskussionsforums polittalk.ch aus den Jahren 2005–2008 besteht und rund 630.000 laufende Wortformen umfasst.

39 Vgl. WEIGT (1968), HOFMEIER (1968), KOPLIN (1968) und VON WEISS (1974).

Typische Nomen

GerMov-Tonbandprotokollekorpus vs. politalk.ch

GerMov-Tonbandprotokollekorpus

Abend **Aktion** Alternative Ansatz Anspruch Arbeiter Aspekt **Assistent**
 Auffassung Auseinandersetzung Basis Bedingung **Bedürfnis** Begriff Beispiel
 Beruf **Betrieb** Bewegung **Beziehung** Brief Brot Bus **Bürokratie** Charakter
Demonstration Ding Diskussion Ebene Eindruck Element **Entwicklung**
 Ereignis **Erfahrung** Erklärung **Fabrik** Fall Faschismus Forderung **Form** Frage
 Frau **Funktion** Gegenstand Genosse Gesellschaft Gespräch Gewalt Grund
Gruppe Herrschaft Hoffnung **Institut** Institution Interesse Justiz Kampf
 Kapitalismus Klasse **Klassenkampf** **Kommune** Kommunikation Kraft
 Kritik Lage **Leute** Logik Lust Mann Mark Maß Minute Modell **Moment**
 Musik Mädchen Möglichkeit Nacht Opposition Ordnung **Organisation** Position
 Praxis Produktion **Professor** Prozeß Punkt Rahmen **Raum** Revolution
Rocker Sache **Scheiße** Schicht SDS Sinn **Situation** Sozialismus
 Sozialist Stand Stein Stellung **Strategie** **Student** Stunde System Tag
Technologie Tendenz Terror Theorie **Universität** Veranstaltung Verbindung
 Verhältnis Vertreter Veränderung Voraussetzung Vorschlag Vorstellung Weg
 Widerstand Wirklichkeit Woche **Wohnung** Wunsch Zeitung Zentrum Zukunft

Abb. 1

5. Kommunikationsstile in korpuslinguistischer Perspektive

5.1 Berechnungsparameter

Da uns aufgrund der Lemmaanalysen über die Annotation von ideologischen Schlagwörtern hinaus keine inhaltlichen Aspekte des Korpus stilrelevant erschienen, haben wir uns dafür entschieden, bei der folgenden komplexen n-Gramm-Analyse auf der Dimension Wortform nur Funktionswörter und Satzzeichen zu berücksichtigen und Lemmata als Dimension gänzlich wegzulassen. Auf der Ebene der Wortarteninformationen gab es hingegen keine Restriktionen. Berechnet wurden Pentagramme ohne Leerstellen, die auch satzübergreifend sein konnten. Die Mindestfrequenz für ein n-Gramm durfte in der Summe beider Korpora nicht kleiner als 4 sein.

5.2 Ergebnis der Analyse des Tonbandprotokollekorpus

Für das Kommunekorpus (hedonistisches Selbstverwirklichungsmilieu) wurden 355.899 Pentagrammtoken berechnet, von denen 307 signifikant waren, für das Linksintellektuellenkorpus (linksintellektuelles Milieu) 286.290, von denen 1.459 signifikant waren. Es ist an dieser Stelle aus Platzgründen nicht möglich, die Analyse vollständig auszuführen. Wir beschränken uns daher auf eine Analyse der auffälligsten Muster und Mustertypen.

<i>Muster</i>	<i>Frequenz Kommunen</i>	<i>Links- intellektuelle</i>	<i>LLR</i>	<i>p</i>
, PDS VVFIN PPER ADV	11	0	12.99	<0.0001
, dann VVFIN PPER ADV	11	0	12.99	<0.0001
, PPER VAFIN ADV ADV	14	1	10.79	<0.0005
\$. ADV , KOUS PPER	10	0	11.80	<0.0003
ADV , PPER VAFIN ADV	10	0	11.80	<0.0003
, ADV VMFIN PPER ADV	10	0	11.80	<0.0003
, PPER VVFIN ADV	10	0	11.80	<0.0003
, , KOUS PPER PRF ADV	10	0	11.80	<0.0003
ADV , PDS VAFIN ADV	9	0	10.62	<0.0005
, ADV VVFIN PPER ADV	19	4	7.64	<0.0031
\$. PPER VVFIN ADV nicht	8	0	9.44	<0.0010
, KON PDS VAFIN ADV	8	0	9.44	<0.0010
ADV VMFIN PPER ADV ADV	8	0	9.44	<0.0010
PDS VVFIN PPER ADV ADV	8	0	9.44	<0.0010
, ich VVFIN ADV , KON PPER VAFIN ADV ADV	8	0	9.44	<0.0010
	8	0	9.44	<0.0010

, PPER VVFIN ADV nicht	8	0	9.44	<0.0010
, KON PDS VAFIN ADV	8	0	9.44	<0.0010
, KON KOUS PPER ADV	13	2	6.80	<0.0051
, PDS VAFIN ADV ADV	13	2	6.80	<0.0051

Tab. 5: Typische sprachliche Muster im Teilkorpus Hedonistisches Selbstverwirklichungsmilieu des GerMov-Tonbandprotokollekorpus

Tabelle 5 zeigt die 20 nicht-inklusive Muster mit den höchsten Signifikanzwerten des Subkorpus Hedonistisches Selbstverwirklichungsmilieu des GerMov-Tonbandprotokollekorpus. Es finden sich auffällig viele Muster, die substituierende Demonstrativpronomen (PDS) enthalten, was auf eine hohe Dichte an anaphorischen Verweisen schließen lässt. Das substituierende Demonstrativpronomen kann gefolgt von einem Vollverb am Anfang eines Gliedsatzes...

, PDS VVFIN PPER ADV	, das weiß ich auch
LLR: 12.99, p: 0.0001	, das weißt du auch
	, das finde ich so
	, das tun wir auch
	, den brauchen wir auf

... oder am Satzanfang stehen ...

\$. PDS VVFIN PPER ADV	. Das finde ich ein
LLR: 6.797, p: 0.0051	. Die kennen wir doch
	. Das hat er schon
	. Das machen wir immer
	? Das weiß ich nicht

... oder gefolgt von einem Hilfsverb am Anfang eines Gliedsatzes ...

, PDS VAFIN ADV ADV	, das war gar nicht
LLR: 6.797, p: 0.0051	, das ist wie so
	, das ist ja wieder
	, das ist ja nicht
	, der ist doch schon

... bzw. nach einer Konjunktion am Anfang eines Satzes.

\$. KON PDS VAFIN ADV	. Und das war sowieso
-----------------------	-----------------------

LLR: 9.44, p: 0.001

- . Und das war also
- . Aber die sind noch
- . Aber das war so
- . Und die war damals

Häufig finden sie sich auch nach vom Tagger als Adverbien klassifizierten Diskurspartikeln, offenbar häufig am Beginn eines Turns.

\$. ADV \$, PDS VAFIN

- . Hier , das ist

LLR: 7.72, p: 0.003

- . Ja , das ist
- . Also , die ist
- . Nein , das ist
- ? Ja , das sind

Im Korpus mit Protokollen des hedonistischen Selbstverwirklichungsmilieus sind also Muster mit anaphorischen Verweisen signifikant, die als ein Indikator für konzeptionell mündlichen Stil gedeutet werden können. Im Vergleich dazu findet sich in der Menge signifikanter Muster im Korpus mit Protokollen des linksintellektuellen Milieus lediglich ein Muster, das ein substituierendes Demonstrativpronomen enthält:

PDS \$, was PPER ADV

- das , was wir bisher

LLR: 6.463, p: 0.0062

- das , was ich vor
- dessen , was Sie bis

Das Pronomen verweist hier nicht auf ein vorher Gesagtes, sondern wird durch einen anschließenden Relativsatz erläutert. Die kontextsensitive Analyse zeigt also, dass die Verwendung substituierender Demonstrativpronomen in diesem Subkorpus eine andere Funktion hat, als im Subkorpus mit Texten des hedonistischen Selbstverwirklichungsmilieus.

Ebenfalls als Indikator für konzeptionelle Mündlichkeit kann die Realisierung einer konditionalen Relation mit der Konjunktion *dann* gelten, die ebenfalls ein signifikantes Mustern des Kommune-Subkorpus ist. In konzeptionell schriftlichen Texten wäre eher zu erwarten, dass die konditionale Relation durch Verberststellung im nachgestellten Hauptsatz realisiert würde. Die Konjunktion *dann* wird sowohl in Verbindung mit Voll- und Hilfsverben, ...

, dann VVFIN PPER ADV

- , dann liegt es doch

LLR: 12.99, p: 0.0001

- , dann hätten wir auch
- , dann gehen wir jetzt
- , dann würden wir quasi
- , dann wird sie vielleicht

... als auch mit Modalverben signifikant häufiger verwendet.

, dann VMFIN PPER ADV	, dann wollte sie gar
LLR: 8.26, p: 0.0021	, dann können sie auch
	, dann können wir ja
	, dann kann ich nur
	, dann kannst du noch

Auffällig sind zudem zahlreiche Muster mit einer hohen ADV-Dichte. Unter ADV subsumiert der TreeTagger Adverbien im engeren Sinn, aber auch Partikeln, insbesondere Modal-, Abtönungs-, Grad- und Fokuspartikeln. Zur Illustration sei das folgende Muster mit einigen seiner Realisierungen im Korpus angeführt:

, PPER VAFIN ADV ADV	, wir sind zwar nicht
LLR: 10.79, p: 0.0005	, es ist doch nur
	, es ist ja auch
	, es wäre doch schon
	, es ist also so

Während sich in den 1.459 signifikanten Mustern des Linksintellektuellen-Subkorpus gerade einmal 20 finden, in denen zwei Adverbien nach einander auftreten, sind es beim Kommunen-Subkorpus rund dreimal so viele bei 307 signifikanten Mustern. Auch die häufige Verwendung von ADV-Clustern kann als Mündlichkeitsindikator gedeutet werden, die häufige Verwendung von Abtönungs- und Modalpartikeln darüber hinaus als Indikatoren eines auf Verständigung abzielenden Kommunikationsstils, in dem allzu klare Festlegungen auf einen Standpunkt vermieden werden. Diese Deutung kann auch mit der Tatsache fundiert werden, dass nur im Kommunekorpus Muster mit Kommunikationsverben wie das folgende signifikant häufig auftreten.

ADV	auch sagen , damit ich
VVINFIN KOMMVERB,	noch sagen , daß es
KOUS PPER	nur sagen , daß du
LLR: 7.08, p: 0.0043	nur sagen , daß sie
	nur sagen , daß ich

Ein weiterer Mündlichkeitsindikator in den Kommuneprotokollen sind Satzanschlüsse mit Konjunktionen, insbesondere mit der Konjunktion *und*. Zwar finden sich hierfür keine Belege in den 20 Mustern mit der höchsten Signifikanz, dennoch lassen sich insgesamt vier Muster wie das Folgende identifizieren:

\$. Und KOUS ich ADV

LLR: 5.9, p: 0.0087

. Und als ich dann

. Und wenn ich jetzt

. Und da ich also

. Und als ich hier

. Und als ich dort

Dass Stilanalysen erst durch den Vergleich an Profil gewinnen, dafür ist Tabelle 6, die die 20 nicht-inklusiven Muster mit den höchsten Signifikanzwerten des Subkorpus linksintellektuelles Milieu des GerMov-Tonbandprotokollekorpus enthält, ein eindrucksvoller Beleg.

<i>Muster</i>	<i>Frequenz Kommunen</i>	<i>Linksintel- lektuelle</i>	<i>LLR</i>	<i>p</i>
NN APPR ART ADJA NN	2	70	97.19	<0.0001
ART NN ART ADJA NN	1	61	89.51	<0.0001
APPR ART NN ART NN	10	65	57.93	<0.0001
ART ADJA NN ART NN	3	49	59.78	<0.0001
ADJA NN APPR ART NN	7	48	43.89	<0.0001
APPR ART NN ART ADJA	0	28	45.24	<0.0001
NN ART NN APPR ART	0	27	43.63	<0.0001
ART NN APPR ART ADJA	1	29	39.27	<0.0001
APPR ART ADJA NN ART	0	26	42.01	<0.0001
ADJA NN ART ADJA NN	1	27	36.18	<0.0001
ADJA NN APPR ART ADJA	1	27	36.18	<0.0001
ART ADJA NN APPR ART	6	35	29.50	<0.0001
ADV APPR ART NN ART	0	22	35.55	<0.0001
ART NN ART NN APPR	2	26	29.96	<0.0001
NN ART NN ART NN	2	25	28.50	<0.0001
ART NN und ART NN	1	22	28.50	<0.0001
APPR ART NN, PRELS	5	30	25.67	<0.0001
ART ADJA NN ART ADJA	0	19	30.70	<0.0001
ART NN ART NN ART	1	21	26.98	<0.0001
, KOUS ART NN ART	4	26	23.17	<0.0001

Tab. 6: Typische sprachliche Muster im Teilkorpus Linksintellektuelles Milieu des GerMov-Tonbandprotokollekorpus

Sie zeigt fast ausschließlich Muster, die Nominal- und Präpositionalphrasencluster repräsentieren. Das Muster mit der höchsten Spezifik repräsentiert eine Folge von Nominal- und Präpositionalgruppe:

NN APPR ART ADJA NN

LLR: 97.19, p: <0.0001

SDS in der ersten Hälfte

Kommilitonen in den anderen Fakultäten
Veränderungen auf einer institutionellen Ebene

Mittelforderung für den autonomen Bereich

Rückfall auf den status quo

Fast immer ist dabei bei einer Folge von Nominalgruppen die zweite durch ein Genitivattribut realisiert, wie das folgende Musters illustriert:

ART ADJA NN ART NN

LLR: 59.87, p: <0.0001

den status quo der Ordinarienprivilegien

der drittelparitätischen Zusammensetzung
der Seminarversammlung

die konkrete Umorganisation des Studiums

die materiellen Ressourcen des Instituts

ein großer Teil der Kommilitonen

Ein weiteres Muster mit hoher Spezifik besteht sogar aus drei aufeinanderfolgenden Präpositional- und/oder Nominalgruppen:

NN ART NN ART NN

LLR: 28.5, p: <0.0001

Form des Studentenproteste der Hochschulrevolte

Willen der Mehrheit der Hochschullehrer

Teilen des SDS das Prinzip

Rest des Semesters die Tätigkeit

Situation der Handelnden den Aktionsmöglichkeiten

Der Grad der Spezifik und die Frequenz dieser Mustertypen verweist auf eine grundlegende stilistische Differenz zwischen den beiden Korpora. Während das Korpus der Kommuneprotokolle durch konzeptionelle Mündlichkeit geprägt ist, scheint das Korpus im linksintellektuellen Milieu eher von Mustern durchwirkt, die gewöhnlich als Indikatoren für Nominalstil beschrieben werden. Hierfür lassen sich weitere Belege anführen.

So finden sich im Korpus mit Protokollen aus dem linksintellektuellen Milieu auch häufiger Muster, die auf eine hohe Frequenz passivischer Formulierungen schließen lassen. Beispielsweise Muster im Perfekt...

NN VVPP VVPP VAFIN
\$.

LLR: 6.46, p: 0.0062

Universität lahmgelegt worden wäre .
Pferdestall angezündet worden ist .
Weg fotografiert worden sind .
Presse rezipiert worden ist ?
Qualität erreicht worden ist .

... oder im Präteritum oder Präsens ...

ADJA NN VVPP VVFIN ,

LLR: 6.62, p: 0.0056

bürgerlichen Presse benutzt werden ,
solche Entwicklungen verhindert werden ,
individuellen Psychoterrors verkündet
wird ,
in Justizfragen mitgearbeitet wurde ,
kleinen Gruppen gemacht wird ,

... oder mit Modalverben:

NN VVPP VVINF VMFIN
\$.

LLR: 14.54, p: <0.0001

Lehrbetrieb ausgesetzt werden muß .
Konsens erzielt werden kann .
Verbindlichkeit geführt werden soll .
Modelle durchgesetzt werden können .
Studentenrat verhandelt werden können .

Schließlich lassen sich im Linksintellektuellenkorpus 18 signifikante Muster mit dem Pronomen *es* und 4 mit dem unpersönlichen Pronomen *man* identifizieren, während das Kommunekorpus lediglich 4 signifikante Muster mit *es* und keines mit *man* enthält.

Die Analyse im Hinblick auf Schlagwörter der Neuen Linken ergibt ebenfalls ein eindeutiges Bild: Für das Kommunekorpus sind keine Muster mit Schlagwörtern stilprägend, im Linksintellektuellenkorpus finden sich hingegen 37 signifikante Muster mit Schlagwort-Nomen ...

NN_SCHLAG APPR ART
NN ART

LLR: 8.07, p: 0.0023

, daß der objektive Prozeß
, ob die objektive Lage
, daß der kommunistische Mensch
, wo die objektiven Bedingungen
, daß die objektive Lage

... oder Adjektiven:

, KOUS ART
ADJA_SCHLAG NN

LLR: 8.07, p: 0.0023

Tendenzen in der Dynamik des
Erfahrung mit der Instituierung eines
Technokraten mit der Differenzierung der
Studenten bei der Durchsetzung der
Widerstand gegen das Vorgehen der

So verdichtet sich für das Korpus mit Protokollen aus dem linksintellektuellen Milieu das Bild von einem nominalen szientistischen Stil⁴⁰, der sich allerdings durch einige Spezifika vom wissenschaftlichen Diskurs unterscheidet.

An erster Stelle ist hier die Verwendung von Kampfvokabular zu nennen. So lassen sich im Linksintellektuellenkorpus 14 signifikante Muster, die Schlagwörter enthalten, auf die Verwendung von Kampfvokabular zurückführen, etwa wie im Fall des folgenden Musters:

ART NN_KAMPF APPR	der Angriff bei der Polizei
ART NN	die Auseinandersetzung auf dem Tegeler
LLR: 16.16, p: <0.0001	den Aktionen vor dem Landgericht
	die Aktion vor dem Landgericht
	den Auseinandersetzungen in der Osnabrücker

Daneben ist eine relativ starke normative Prägung der Protokolle zu erkennen, wofür schon die Zahl von signifikanten Mustern mit Modalverben ein Indikator ist. Eine genauere Prüfung der Muster und ihrer Repräsentanten bestätigt diese Annahme. Solche Muster sind etwa Satzanfänge mit Personalpronomen und Modalverb ...

VAFIN . PPER VMFIN	ist . Wir dürfen nicht
ADV	sind . Wir müssen also
LLR: 8.08, p: 0,024	sind . Wir müssen dahin
	war . Wir müssen jetzt
	ist . Es könnte aber

... oder Konstatierungen eines Imperativs mit *müssen* nach einem Heckenausdruck:

Ich VVFIN , PPER	Ich finde , wir müssen
VMFIN	Ich meine , wir müssen
LLR: 6.46, p: 0.006	Ich glaube , wir müssen

Insgesamt zeigt die Analyse der Ergebnisse des datengeleiteten Korpusvergleichs erhebliche Differenzen zwischen den beiden untersuchten Subkorpora des GerMov-Korpus, die auch insofern als stilistische Unterschiede interpretiert werden können, als sich unterschiedliche Mustertypen mehr oder weniger kohärent unter ein Deutungsschema (etwa Nominalstil, konzeptionelle Mündlichkeit) subsumieren ließen.

40 Vgl. BRAUN (1979: 53), PUSCH (1976: 17ff.), KAEWWIPAT (2001).

5.3 Validierung durch das Flugblattkorpus

Können die rekonstruierten Stile nun auch im Flugblattkorpus nachgewiesen werden und lassen sich die Untersuchungsergebnisse so validieren? Zur Beantwortung dieser Fragen haben wir analog zur komplexen n-Gramm-Analyse zum Tonbandprotokollekorpus eine Analyse des nach Milieus gefilterten Flugblattkorpus durchgeführt. Für das Kommunekorpus (hedonistisches Selbstverwirklichungsmilieu) wurden 114.090 Pentagrammtoken berechnet, von denen 351 signifikant waren, für das Linksintellektuellenkorpus (linksintellektuelles Milieu) 384.212, von denen 175 signifikant waren.⁴¹

Im Anschluss an die n-Gramm-Analyse haben wir berechnet, wie viel Prozent der signifikanten Muster sich auch in der Analyse zum Tonbandprotokollekorpus als signifikant für das jeweilige Korpus erwiesen hatten. Dabei wurde die jeweils kleinste Menge typischer n-Gramme als Berechnungsbasis gewählt. Während die Korpora des linksintellektuellen Milieus 16 % Übereinstimmung im Hinblick auf signifikante Muster zeigten (28 gemeinsame Muster), waren es gerade einmal zwei Muster (0,007 %), die beim hedonistischen Selbstverwirklichungsmilieu übereinstimmten. Die konzeptionelle Mündlichkeit blieb beim Wechsel in ein anderes Medium nicht in gleicher Weise erhalten. Betrachtet man hingegen die im Folgenden aufgelisteten übereinstimmenden Muster von Flugblatt- und Tonbandprotokollekorpus aus dem linksintellektuellen Milieu, so zeigt sich, dass die Indikatoren für Nominalstil im Mündlichen wie im Schriftlichen stilprägend sind:

41 Wie die im Vergleich zur Tonbandprotokollekorpus-Analyse geringere Anzahl signifikanter Muster zeigt, sind die stilistischen Unterschiede beim Flugblattkorpus nicht so prägnant. Allerdings ist die geringe Zahl signifikanter Muster auch der geringen Korpusgröße, insbesondere bei den Flugblättern aus dem hedonistischen Selbstverwirklichungsmilieu, geschuldet.

ADJA NN APPR ART ADJA
 ADJA NN APPR ART NN
 ADJA NN ART ADJA NN
 APPR ART ADJA ADJA NN
 APPR ART ADJA NN APPR
 APPR ART ADJA NN KON
 APPR ART NN APPR NN
 APPR ART NN ART NN
 APPR ART NN von NN
 APPR NN APPR ART NN
 ART ADJA KON ADJA NN
 ART ADJA NN APPR ART
 ART ADJA NN APPR NN
 ART ADJA NN ART NN
 ART NN APPR ART ADJA
 ART NN ART ADJA NN
 ART NN ART NN \$,
 ART NN ART NN ,
 ART NN ART NN APPR
 ART NN ART NN ART
 NN \$, PRELS ART NN
 NN , PRELS ART NN
 NN APPR ART ADJA NN
 NN APPR ART NN ART
 NN ART ADJA NN ART
 NN ART NN ART ADJA
 NN ART NN ART NN
 in ART NN ART NN

Eine genauere Analyse jener Muster, die im Flugblattkorpus für das hedonistische Selbstverwirklichungsmilieu signifikant sind, lässt aber auch hier Indikatoren erkennen, die sich im Sinne der rekonstruierten stilistischen Dimension der konzeptionellen Mündlichkeit deuten: kurze Sätze, Vulgarismen, das Duzen der Leserinnen und Leser etc.

6. Fazit

Die exemplarische Analyse hat gezeigt, dass die von unserer Forschergruppe *semtracks* entwickelte Methode der datengeleiteten Stilanalyse eine geeignete Methode zur Rekonstruktion kommunikativer Stile sein kann. Zwar war die Zahl der signifikanten Muster gerade noch überschaubar, bei größeren Korpora allerdings und bei anders gewählten Berechnungsparametern (z.B. Tetragramme mit Einbeziehung der Varianten auf allen Dimensionen in einem Fenster von vier bis acht token) können leicht unübersichtliche Mengen an signifikanten Mustern zusammenkommen, die kaum noch einer kohärenten Auswertung unterzogen werden können. So besteht ein wesentliches Desiderat der Methode in einer intelligenten Zusammenfassung der Daten, vorzugsweise in einer automatisierten Gruppierung funktional ähnlicher Muster.

Dennoch zeigen die Ergebnisse auch, dass der datengeleitete Korpusvergleich von Korpora, die in möglichst wenigen außersprachlichen Variablen variieren, das Potenzial hat, zu einer zentralen Methode der Korpuspragmatik zu werden. Die komplexe n-Gramm-Analyse ist hier freilich nicht die einzige Methode, sprachliche Muster mit pragmatischer Ladung zu identifizieren. Die Ausarbeitung eines differenzierten Methodenrepertoires und die Verbesserung der Validität und Reliabilität ihrer Messinstrumente wird die Agenda der Korpuspragmatik in den nächsten Jahren bestimmen.

Literatur

- BELICA, CYRIL (1996): Analysis of temporal changes in corpora. In: *International journal of corpus linguistics* 1.1 (1996). 61–73.
- BIEDERMANN, REINHARD (1969): Die deutschen Gradadverbien. Heidelberg: [Dissertation] Universität Heidelberg.
- BIERWISCH, MANFRED (1987): Semantik der Graduierung. Grammatische und konzeptuelle Aspekte von Dimensionsadjektiven. In: BIERWISCH, MANFRED / LANG, EWALD (Hgg.): Grammatische und konzeptuelle Aspekte von Dimensionsadjektiven. Berlin: Akademie (Studia Grammatica 26/27). 91–286.
- BOURDIEU, PIERRE (1993): Sozialer Sinn. Kritik der theoretischen Vernunft. Frankfurt am Main.

- BRAUN, PETER (1979): Tendenzen in der deutschen Gegenwartssprache. Stuttgart.
- BUBENHOFER, NOAH (2006): Einführung in die Korpuslinguistik. Praktische Grundlagen und Werkzeuge. Elektronische Ressource.
URL: www.bubenhofer.com/korpuslinguistik/
- BUBENHOFER, NOAH (2009): Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. Berlin (Sprache und Wissen 4).
- BUBENHOFER, NOAH / DUSSA, TOBIAS / EBLING, SARAH / KLIMKE, MARTIN / ROTHENHÄUSLER, KLAUS / SCHARLOTH, JOACHIM / TAMEKUE, SUARÈS / VOLA, SASKIA (Forschergruppe *semtracks*) (2009): „So etwas wie eine Botschaft“. Korpuslinguistische Analysen der Bundestagswahl 2009. In: Sprachreport 25.4 (2009). 2–10.
- BUBENHOFER, NOAH / KLIMKE, MARTIN / SCHARLOTH, JOACHIM (2009): Political tracker – Bundestagswahl '09. Eine Semantische Matrixanalyse. Elektronische Ressource. URL: <http://semtracks.com/politicaltracker/>
- BUBENHOFER, NOAH / SCHARLOTH, JOACHIM (2010): Kontext korpuslinguistisch. Die induktive Berechnung von Sprachgebrauchsmustern in großen Textkorpora. In: KLOTZ, PETER / PORTMANN-TSELIKAS, PAUL R. / WEIDACHER, GEORG (Hgg.): Kontexte und Texte. Soziokulturelle Konstellationen literalen Handelns. Tübingen. 85–108 (Europäische Studien zur Textlinguistik 8).
- DORNSEIFF, FRANZ (2004): Der deutsche Wortschatz nach Sachgruppen. Berlin, New York.
- EBLING, SARAH / SCHARLOTH, JOACHIM / DUSSA, TOBIAS / BUBENHOFER, NOAH (2011): Gibt es eine Sprache des politischen Extremismus? In: LIEDTKE, FRANK (Hg.): Die da oben.... Politik und Partizipation im Wahljahr 2009. (in Druck)
- FAHLENBRACH, KATHRIN (2002): Protest-Inszenierungen. Visuelle Kommunikation und Kollektive Identitäten in Protestbewegungen. Opladen.
- FAHLENBRACH, KATHRIN (2007): Protestinszenierungen. Die Studentenbewegung von 1968 im Spannungsfeld von Kultur-Revolution und Medien-Evolution. In: KLIMKE, MARTIN / SCHARLOTH, JOACHIM (Hgg.): 1968. Ein Handbuch zur Kultur- und Mediengeschichte der Studentenbewegung. Stuttgart. 11–23.
- FEILKE, HELMUTH (2000): Die pragmatische Wende in der Textlinguistik. In: BRINKER, KLAUS et al. (Hgg.): HSK Text- und Gesprächslinguistik. Berlin, New York. 64–82.
- FEILKE, HELMUTH / LINKE, ANGELIKA (Hgg.) (2009): Oberfläche und Performanz. Untersuchungen zur Sprache als dynamischer Gestalt. Berlin, New York.
- FIX, ULLA (1987): „Kommunikativ adäquat“– „stilistisch adäquat“. Zu Problemen, Kategorien und Kriterien der Redebewertung. Diss. B. Leipzig.
- FIX, ULLA (1991): Stilistische Textanalyse – immer ein Vergleich? Das Gemeinsame von Methoden der Stilanalyse – das Gemeinsame an Stilbegriffen. In: BRINKER, KLAUS (Hg.): Aspekte der Textlinguistik, Hildesheim, Zürich, New York. 133–156.
- FIX, ULLA (1996): Gestalt und Gestalten. Von der Notwendigkeit der Gestaltkategorie für eine das Ästhetische berücksichtigende pragmatische Stilistik. In: Zeitschrift für Germanistik 2 (1996). 308–323.
- FIX, ULLA (2005): Die stilistische Einheit von Texten – auch ein Textualitätskriterium? In: REUTER, EWALD / SORVALI, TINA (Hgg.): Satz – Text – Kulturkontrast. Frankfurt am Main u. a. 35–50.

- FIX, ULLA (2009): Muster und Abweichen. In: FIX, ULLA / GARDT, ANDREAS / KNAPE, JOACHIM: *Rhetorik und Stilistik/Rhetoric and Stylistics*. Berlin, New York. 1300–1315 (Handbücher zur Sprach- und Kommunikationswissenschaft 31).
- GERMOV: Korpus zur gesprochenen und geschriebenen Sprache der 1968er-Bewegung in der Bundesrepublik Deutschland: Flugblätter und Tonbandprotokolle. Zusammengestellt und herausgegeben von Joachim Scharloth und Noah Bubenhofer unter Mitarbeit von Mechthild Scharloth. (cosmov, Korpora zur Erforschung sozialer Bewegungen 1) Online: <http://www.semtracks.com/cosmov/>
- GROB, MARION (1985): *Das Kleidungsverhalten jugendlicher Protestgruppen in Deutschland im 20. Jahrhundert*. Münster.
- HARRAS, GISELA / WINKLER, EDELTRAUD / ERB, SABINE / PROOST, KRISTEL (Hgg.) (2004): *Handbuch deutscher Kommunikationsverben*. Berlin, New York (Schriften des Instituts für deutsche Sprache).
- HANDLER, PETER (2009): Stilistische Phänomene der Wortbildung. In: FIX, ULLA / GARDT, ANDREAS / KNAPE, JOACHIM: *Rhetorik und Stilistik/Rhetoric and Stylistics*. Berlin, New York. 1563–1575 (Handbücher zur Sprach- und Kommunikationswissenschaft 31).
- HEINEMANN, WOLFGANG (2009): Stilistische Phänomene auf der Ebene des Textes. FIX, ULLA / GARDT, ANDREAS / KNAPE, JOACHIM: *Rhetorik und Stilistik/Rhetoric and Stylistics*. Berlin, New York. 1610–1630.
- HINNENKAMP, VOLKER / SELTING, MARGRET (Hgg.) (1989): *Stil und Stilisierung. Arbeiten zur interpretativen Soziolinguistik*. Tübingen.
- HOCKETT, CHARLES F. (1958): *A Course in Modern Linguistics*. New York.
- HOOVER, DAVID L. (2007): Corpus Stylistics, Stylometry, and the Styles of Henry James. In: *Style* 41.2 (2007). 174–203.
- HOFMEIER, KLAUS (1968): *Lieben Sie Establishment?* Köln.
- HRADIL, STEFAN (2001): *Soziale Ungleichheit in Deutschland*. Opladen.
- KAEWIPAT, NORASETH (2001): Nominalstil im Gegenwartsdeutschen. In: TDLV-Forum. *Zeitschrift für Sprache, Literatur und Kultur* 6 (2001). 21–39.
- KOPLIN, RAIMUND (1968): *Sprachführer durch die Revolution*. München.
- LABOV, WILLIAM (1972): *Sociolinguistic Patterns*. Philadelphia.
- MANNING, CHRISTOPHER D. / SCHÜTZE, HINRICH (2002): *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts.
- MCENERY, TONY / XIAO, RICHARD / TONO, YUKIO (2006): *Corpus-Based Language Studies: An Advanced Resource Book*. New York.
- NASSEHI, ARMIN (1998): Multikulturalität und Knappheit. In: HILLEBRANDT, FRANK / KNEER, GEORG / KRAEMER, KLAUS (Hgg.): *Verlust der Sicherheit? Lebensstile zwischen Multioptionalität und Knappheit*. Opladen. 52–63.
- OS, CHARLES VAN (1989): *Aspekte der Intensivierung im Deutschen*. Tübingen (Studien zur deutschen Grammatik 37).
- PERKUHN, RAINER / BELICA, CYRIL (2006): Korpuslinguistik – Das unbekannte Wesen. Oder: Mythen über Korpora und Korpuslinguistik. In: *Sprachreport* 22.1 (2006). 2–8.
- PUSCH, LUISE F. (1976): Nominalisierungen in der deutschen Sprache der Gegenwart. In: *Jahrbuch Deutsch als Fremdsprache* 1976. 14–55.

- SANDIG, BARBARA (1995): Tendenzen der linguistischen Stilforschung. In: STICKEL, GERHARD (Hg.): *Stilfragen*. Berlin, New York. 27–61.
- SCHARLOTH, JOACHIM (2007): Die Sprache der Revolte. Linke Wörter und avantgardistische Kommunikationsstile. In: KLIMKE, MARTIN / SCHARLOTH, JOACHIM (Hgg.): 1968. *Handbuch zur Kultur- und Mediengeschichte der Studentenbewegung*. Stuttgart. 223–234.
- SCHARLOTH, JOACHIM (2011): 1968. Eine Kommunikationsgeschichte. Paderborn.
- SCHILLER, ANNE / TEUFEL, SIMONE / THIELEN, CHRISTINE (1995): Guidelines für das Tagging deutscher Textcorpora mit STTS. Stuttgart: [Working Paper] Universität Stuttgart, Institut für maschinelle Sprachverarbeitung, Tübingen: Seminar für Sprachwissenschaft.
- SCHMID, HELMUT (1994): Probabilistic part-of-speech tagging using decision trees. Stuttgart: [Working paper] Universität Stuttgart, Institut für maschinelle Sprachverarbeitung.
- SCHULZE, GERHARD (1996): *Die Erlebnisgesellschaft. Kultursoziologie der Gegenwart*. Frankfurt am Main.
- SELTING, MARGRET (1997): Interaktionale Stilistik: Methodologische Aspekte der Analyse von Sprechstilen. In: SELTING, MARGRET / SANDIG, BARBARA (Hgg.): *Sprech- und Gesprächsstile*. Berlin, New York. 9–43.
- SELTING, MARGRET / HINNENKAMP, VOLKER (1989): Einleitung: Stil und Stilisierung in der Interpretativen Soziolinguistik. In: HINNENKAMP, VOLKER / SELTING, MARGRET (Hgg.): *Stil und Stilisierung. Arbeiten zur interpretativen Soziolinguistik*. Tübingen. 1–26 (Linguistische Arbeiten 235).
- SEMINO, ELENA / SHORT, MICK (2004): *Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Writing*. London.
- SINCLAIR, JOHN (1991): *Corpus, concordance, collocation*. Oxford.
- SPILLNER, BERND (2009): Verfahren stilistischer Textanalyse. In: FIX, ULLA / GARDT, ANDREAS / KNAPE, JOACHIM: *Rhetorik und Stilistik/Rhetoric and Stylistics*. Berlin, New York. 1739–1782 (*Handbücher zur Sprach- und Kommunikationswissenschaft* 31).
- STEYER, KATHRIN (2004): Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikografische Perspektiven. In: DIES. (Hg.): *Wortverbindungen – mehr oder weniger fest*. Berlin. 87–116. (Institut für Deutsche Sprache. Jahrbuch 2003)
- STEYER, KATHRIN / BRUNNER, ANNELEN (2009): Das UWV-Analysemodell. Eine korpusgesteuerte Methode zur linguistischen Systematisierung von Wortverbindungen. Mannheim: Institut für Deutsche Sprache (Online publizierte Arbeiten zur Linguistik OPAL 1). Elektronische Ressource. URL: <http://www.ids-mannheim.de/pub/laufend/opal/privat/opal09-1.html>
- STEYER, KATHRIN / LAUER, MEIKE (2007): Corpus-Driven. Linguistische Interpretation von Kookkurrenzbeziehungen. In: EICHINGER, LUDWIG M. / KÄMPER, HEIDRUN (Hgg.): *Sprach-Perspektiven. Germanistische Linguistik und das Institut für Deutsche Sprache*. Tübingen. 493–509 (*Studien zur deutschen Sprache* 40).
- TEUBERT, WOLFGANG (2005): My version of corpus linguistics. In: *International journal of corpus linguistics* 10.1 (2005). 1–13.
- TOGNINI-BONELLI, ELENA (2001): *Corpus linguistics at work*. Amsterdam. (*Studies in corpus linguistics* 6).

- WEIGT, PETER (1968): Revolutions-Lexikon. Taschenbuch der ausserparlamentarischen Aktion. Frankfurt am Main.
- WEISS, ANDREAS VON (1974): Schlagwörter der neuen Linken: die Agitation der Sozialrevolutionäre. München.